

Preplanned Studies

The Potential Distribution Prediction of *Oncomelania hupensis* Based on Newly Emerging and Reemergent Habitats — China, 2015–2019

Yanfeng Gong¹; Yinlong Li^{1,†}; Lijuan Zhang¹; Shan Lv¹; Jing Xu¹; Shizhu Li^{1,†}

Summary

What is already known about this topic?

Oncomelania hupensis (*O. hupensis*) is the only intermediate host of *Schistosoma japonicum*. The distribution of *O. hupensis* is affected by a series of climate, geographical, and socioeconomic factors, which reflect the risk areas of schistosomiasis.

What is added by this report?

There were certain geographical aggregations in the potential high distribution areas of *O. hupensis*, which were mainly distributed in Poyang Lake area, Dongting Lake area, and the middle and the lower reaches of the Yangtze River.

What are the implications for public health practice?

Monitoring and forecasting the distribution of *O. hupensis* is conducive to improving the early warning capabilities of the potential risk of schistosomiasis transmission and progressively promoting the elimination of schistosomiasis nationwide.

The infection rate of schistosomiasis in China now is at the lowest level in history, and the source of infection has been effectively controlled. However, the area of *Oncomelania hupensis* (*O. hupensis*), the only intermediate host of the genus *Schistosoma*, has been maintained at about 3.5 billion m². Affected by the changes in climate, environment, and human activities, newly emerging and reemergent habitats of *O. hupensis* often occur, which makes early warning of *O. hupensis* become particularly important. Based on the surveillance data of *O. hupensis* (1), a niche model was used to predict the distribution of *O. hupensis*. The results indicated that the distribution of *O. hupensis* had concentrated, becoming primarily distributed in the middle and lower reaches of the Yangtze River, Dongting Lake, and Poyang Lake. Forecasting the distribution of *O. hupensis* is conducive to improving the ability of early warning of potential transmission

risk and the elimination of schistosomiasis.

Schistosomiasis, a widespread zoonotic disease transmitted by the parasite of *Schistosoma japonicum* (*S. japonicum*), is considered as one of the most severe public health threats in China. *O. hupensis* is the only intermediate host of *S. japonicum* and plays a key role in schistosomiasis transmission, whose geographical distribution reflects the area at risk of schistosomiasis transmission. According to the National Schistosomiasis Reporting System, the reported snail distribution has been maintained at about 3.5 billion square meters since 2004, and the occurrence and recurrence of snails have been reported in many areas (2). In 2016 and 2020, the Yangtze River Basin and its southern region suffered from 2 catastrophic floods. After the floods, suitable areas for snail breeding were formed, which increased the potential scope and degree of schistosomiasis distribution (3).

The environment with newly emerging and reemergent snail habitats from the national schistosomiasis surveillance sites from 2015 to 2019 were selected as the sample points, and 14 climatic, geographical, and socioeconomic factors (Supplementary Table S1 available in <http://weekly.chinacdc.cn/>) influencing *O. hupensis* distribution were collected as predictors. Spatial scanning based on Poisson model in SaTScan (version 9.6) was used to analyze the aggregation of newly emerging and reemergent habitats of *O. hupensis*. A total of 10 common niche models (Supplementary Table S2 available in <http://weekly.chinacdc.cn/>) provided in biomod2 platform (version 3.4.6) in R software (version 3.6.1; RStudio Inc; the USA) were used to construct single models, and the area under the curve (AUC) and true skill statistics (TSS) were used as the indicators of model screening, which were threshold-related indicators and threshold independent indicators commonly used to evaluate model accuracy. Single models were selected according to the criteria of AUC >0.90 and TSS >0.85. The weights were determined

according to the TSS of the results of different algorithms, and the combined niche model was calculated by the weighted average method. The judgment threshold is defined as the minimum existence threshold, and the distribution probability from 0.00 to the minimum existence threshold is taken as the no-risk area, the minimum existence threshold to 0.70 is the low-risk area, 0.70 to 0.90 is the medium-risk area, and the greater than 0.9 is the high-risk area (4).

According to the results of SaTScan scanning, six cluster areas in newly emerging and reemergent habitats of *O. hupensis* were detected, which were primarily distributed in the northwest region of Yunnan Province, the central part of Sichuan Province, and the junction of Jiangsu Province, Anhui Province, and Zhejiang Province (Figure 1). The prediction performance of 10 niche models was statistically significant through Kruskal-Wallis H test (AUC, $H=24.720$, $P<0.05$; TSS, $H=29.372$, $P<0.05$), and random forest (AUC=0.973, TSS=0.998) demonstrated the best prediction performance (Supplementary Table S3 available in <http://weekly.chinacdc.cn/>). Taken together, the importance of environmental variables in each model is different, of

which the average annual rainfall and average annual temperature had more impacts (Supplementary Table S4 available in <http://weekly.chinacdc.cn/>). The current distribution of *O. hupensis* predicted by combination model indicated that the potential distribution areas of *O. hupensis* were mainly concentrated in the middle and lower reaches of the Yangtze River (Figure 2). These areas were also locally concentrated in the central part of Sichuan Province and scattered in the northwest part of Yunnan Province (Figure 2). Based on the predicted results of the niche model, the potential distribution area of *O. hupensis* accounted for 22.92% of the total area in China, and the distribution area can be categorized into areas of low-distribution (8.74%), medium-distribution (7.99%), and high-distribution (6.19%).

DISCUSSION

The results of this study indicated that the certain geographical aggregations in the potential high distribution areas of *O. hupensis* were primarily distributed in Poyang Lake area, Dongting Lake area, and the middle and the lower reaches of the Yangtze River. These areas maintain high water levels due to

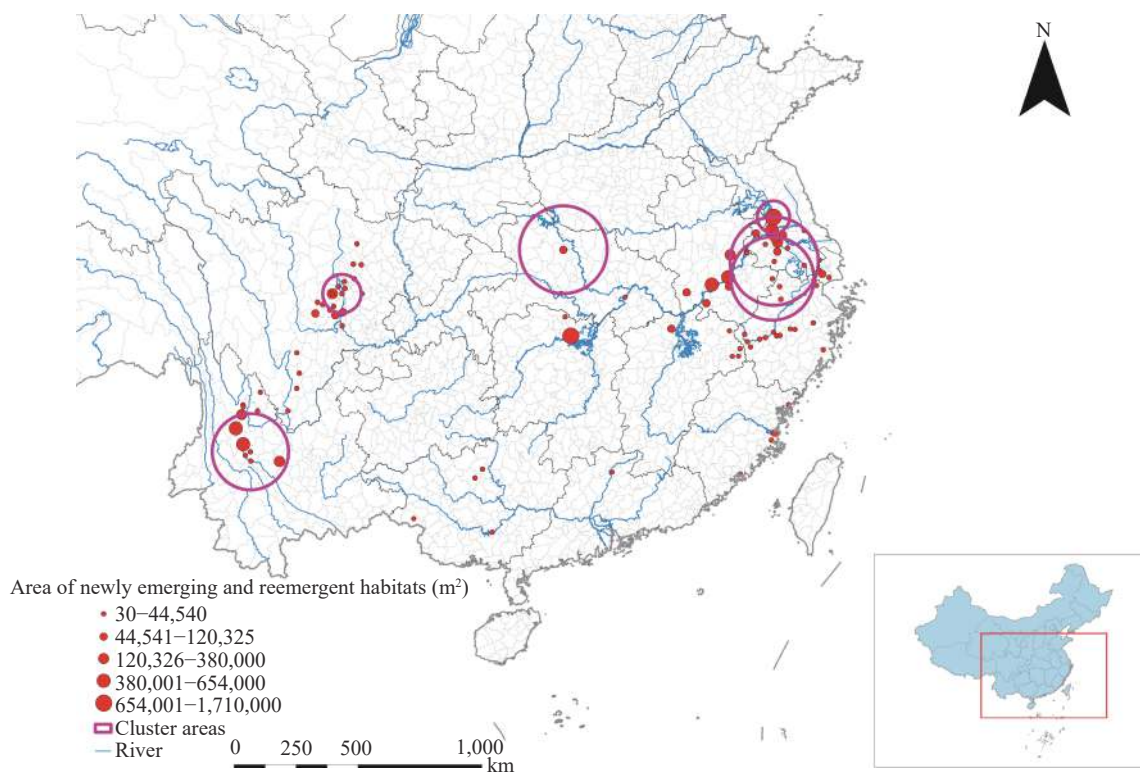


FIGURE 1. Ranges of newly emerging and reemergent habitats of *Oncomelania hupensis* and cluster area in schistosomiasis endemic areas, 2015–2019.

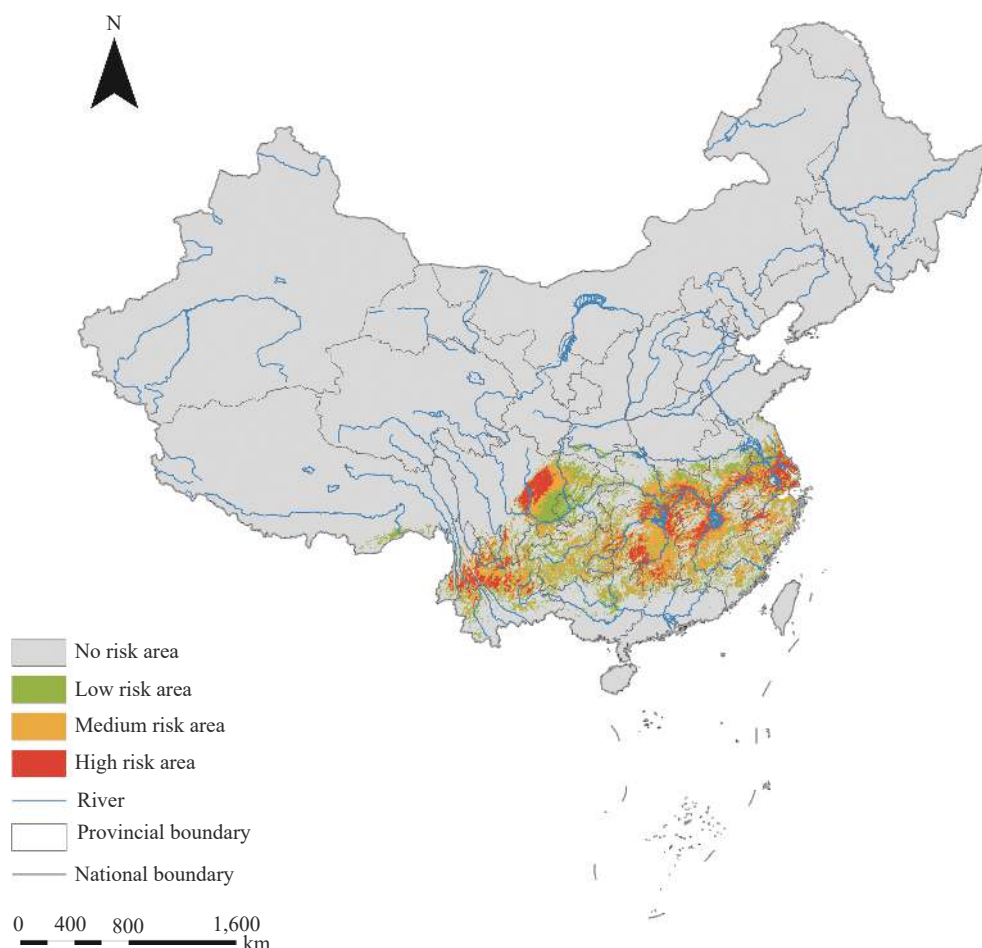


FIGURE 2. Current distribution prediction map of *Oncomelania hupensis* in China based on combination model.

floods in summer and begin to recede in fall and winter. The seasonal periodic water-level changes lead the coastal beaches to form a wet meadow ecosystem, providing a suitable breeding place for *O. hupensis* (5). Due to fishing, grazing, farming, and other production and living reasons, residents were more likely to be exposed to snail environment and contaminated water (6). *O. hupensis* was distributed locally in the central part of Sichuan Province and scattered in the northwest part of Yunnan Province. Different from lake-marsh type snails distributed along the beach, the general law of the distribution of hilly snails is that they are distributed in the breeding environment of different topography with the water system from top to bottom. The migration and diffusion of *Oncomelania* snails are related to the size, distribution, velocity, and irrigation of water system. Affected by flood erosion, *O. hupensis* with higher terrain may spread with water flowing down the mountain, increasing the potential risk of schistosomiasis transmission (7).

Compared with the research of Liao et al. (8), this

study used the latest snail survey data and combined the niche model to overcome the uncertainty brought by the single model. The high-risk areas of *O. hupensis* are basically consistent with the distribution patterns of schistosomiasis transmission in China (9). Compared with the risk distribution of schistosomiasis transmission assessed by Duan et al. (10), the high-risk areas predicted in this study covered a wider area, while the range of middle-areas and low-risk areas was lower than predicted and might be related to different sample points and predictors of two studies.

This study was subjected to some limitations. First, newly emerging and reemergent habitats outside the surveillance sites might have occurred, and this study did not take these sample points into account. Second, due to the unavailability of data, this study did not consider the impact of other human factors such as water conservancy projects, water supply reconstruction, and lavatories. In future research, more complete data of snail distribution are essential and can be accurately obtained through field investigation.

More relevant variables from socioeconomic and environmental fields are necessary to be added to model training to improve the accuracy and effectiveness of the model prediction, promote the prediction results of potential snail distribution areas, and provide further technical support for the prevention, control, and monitoring of schistosomiasis.

Affected by production activities, natural disasters, water conservancy projects, climate change, and other factors, the abnormal spread of *O. hupensis* leads to the recurrence of schistosomiasis in some areas and even the spread to historically non-endemic areas. Historical experience suggests that there will be a relative peak in the area and the density of snails three years after flooding, which will increase the risk of repeated outbreaks of schistosomiasis and difficulty in schistosomiasis control (6). Although the national schistosomiasis control program in China has made great strides towards the interruption of transmission, the risk of schistosomiasis transmission still exists in some areas. To further consolidate the results of the national schistosomiasis control program and prevent rebound of schistosomiasis, taking comprehensive control strategies and measures in areas at risk of *O. hupensis* infestation is necessary.

Conflicts of interest: There are no conflicts of interest to declare.

doi: [10.46234/ccdcw2021.023](https://doi.org/10.46234/ccdcw2021.023)

* Corresponding authors: Yinlong Li, liyl@nprd.chinacdc.cn; Shizhu Li, lisz@chinacdc.cn.

¹ National Institute of Parasitic Diseases, Chinese Center for Disease Control and Prevention; Chinese Center for Tropical Diseases Research; WHO Collaborating Centre for Tropical Diseases; National Center for International Research on Tropical Diseases, Ministry of Science and Technology; Key Laboratory of Parasite and Vector

Biology, National Health Commission of China, Shanghai, China.

Submitted: December 07, 2020; Accepted: January 27, 2021

REFERENCES

- Xu J, Li SZ, Zhang LJ, Bergquist R, Dang H, Wang Q, et al. Surveillance-based evidence: elimination of schistosomiasis as a public health problem in the People's Republic of China. *Infect Dis Poverty* 2020;9(1):63. <http://dx.doi.org/10.1186/s40249-020-00676-5>.
- Cao CL, Zhang LJ, Bao ZP, Dai SM, Lü S, Xu J, et al. Endemic situation of schistosomiasis in People's Republic of China from 2010 to 2017. *Chin J Schisto Control* 2019;31(5):519 – 21,554. <http://dx.doi.org/10.16250/j.32.1374.2018232>. (In Chinese).
- Wang H, Xiong YL, Zhang JJ, Li Y, Zuo YT, Xu MX. Assessment of schistosomiasis transmission risk after flood damage in Wuhan City. *Chin J Schisto Control* 2018;30(4):410 – 4. <http://dx.doi.org/10.16250/j.32.1374.2018046>. (In Chinese).
- Hu XK, Hao YW, Xia S, Guo YH, Xue JB, Zhang Y, et al. Detection of schistosomiasis transmission risks in Yunnan Province based on ecological niche modeling. *Chin J Parasitol Parasit Dis* 2020;38(1):80 – 6,94. <http://dx.doi.org/10.12140/j.issn.1000-7423.2020.01.012>. (In Chinese).
- Zou L, Ruan SG. Schistosomiasis transmission and control in China. *Acta Trop* 2015;143:51 – 7. <http://dx.doi.org/10.1016/j.actatropica.2014.12.004>.
- Cao CL, Li SZ, Zhou XN. Impact of schistosomiasis transmission by catastrophic flood damage and emergency response in China. *Chin J Schisto Control* 2016;28(6):618 – 23. <http://dx.doi.org/10.16250/j.32.1374.2016215>. (In Chinese).
- Lv SB, Lin DD. Natural environment and schistosomiasis transmission in Poyang Lake region. *Chin J Schisto Control* 2014(5):561 – 4,574. <http://dx.doi.org/10.3969/j.issn.1005-6661.2014.05.024>. (In Chinese).
- Liao JS. Study on the potential transmission risks of schistosomiasis japonica in China based on ecological niche models [dissertation]. Hunan: Central South University; 2011. <http://dx.doi.org/10.7666/d.y1914579>. (In Chinese).
- Shi L, Li W, Wu F, Zhang JF, Yang K, Zhou XN. Epidemiological features and control progress of schistosomiasis in waterway-network region in the People's Republic of China. *Adv Parasitol* 2016;92:97 – 116. <http://dx.doi.org/10.1016/bs.apar.2016.02.002>.
- Duan JQ, Zhou J, Huang DP, Liu LL. Transmission risk assessment of schistosomiasis in China based on maximum entropy model. *Meteorol Sci Technol* 2019;47(3):520 – 5. <http://dx.doi.org/10.19517/j.1671-6345.20180347>. (In Chinese).

SUPPLEMENTARY TABLE S1. Summary of environment variables used in the study.

Category	Variable name	Source
Climate factor	≥0 °C annual accumulated temperature	http://www.resdc.cn
	≥10 °C annual accumulated temperature	
	Aridity	
	Index of Moisture	
	Average annual precipitation	
	Average annual temperature	
Geographical factor	Landform	http://www.geodata.cn
	Land use	
	Elevation	http://www.resdc.cn
	Sand	
	Silt	
Socioeconomic factor	Clay	http://www.resdc.cn
	Gross domestic product	
	Density of population	

SUPPLEMENTARY TABLE S2. Summary of niche models used in the study.

Classification	Model name
Environmental envelope theory	Surface range envelope
Statistical regression algorithm	Generalized linear model
	Generalized additive model
	Multivariate adaptive regression spline
Statistical classification algorithm	Generalized boosted model
	Classification tree analysis
	Flexible discriminant analysis
Machine learning algorithm	Artificial neural networks
	Random forest
	Maximum entropy

SUPPLEMENTARY TABLE S3. Performance evaluation indicators results of 10 niche models used in the study.

Model	TSS (mean ± SD)	AUC (mean ± SD)
GLM	0.868 ± 0.057	0.947 ± 0.042
GBM	0.950 ± 0.007	0.996 ± 0.001
GAM	0.926 ± 0.008	0.990 ± 0.001
CTA	0.945 ± 0.009	0.980 ± 0.005
ANN	0.813 ± 0.056	0.930 ± 0.032
SRE	0.651 ± 0.018	0.825 ± 0.008
MARS	0.846 ± 0.251	0.946 ± 0.133
FDA	0.914 ± 0.015	0.983 ± 0.006
RF	0.973 ± 0.004	0.998 ± 0.001
MaxEnt	0.815 ± 0.023	0.915 ± 0.023

Abbreviation: TSS=True skill statistics; AUC=Area under the curve; SD=Standard deviation; GLM=Generalized linear model; GBM=Generalized boosted model; GAM=Generalized additive model; CTA=Classification tree analysis; ANN=Artificial neural networks; SRE=Surface range envelope; MARS=Multivariate adaptive regression spline; FDA=Flexible discriminant analysis; RF=Random forest; MaxEnt=Maximum entropy.

SUPPLEMENTARY TABLE S4. Importance percentage (%) of environmental variables of 10 niche models used in the study.

Model	GLM	GBM	GAM	CTA	ANN	SRE	MARS	FDA	RF	MaxEnt
AAT	22.02	34.93	29.21	35.22	13.02	11.88	43.54	25.83	27.62	19.01
AAP	17.70	18.61	15.72	24.72	16.44	11.54	12.57	26.11	10.35	5.75
IM	18.45	9.73	8.24	0.34	31.16	10.03	17.78	15.69	4.75	9.11
LF	7.28	24.28	13.29	12.15	6.78	10.68	5.31	8.33	15.85	7.13
GDP	1.07	5.62	0.03	15.76	2.16	9.71	2.57	8.40	16.49	9.36
AR	6.32	1.22	5.53	0.21	11.17	9.17	5.71	2.92	0.90	11.02
EL	10.76	0.03	2.07	0.60	10.73	4.75	0.88	1.77	4.75	9.04
AAT0	8.61	0.48	5.56	2.53	0.53	6.39	2.79	6.08	10.29	9.18
DP	1.82	0.45	0.22	0.12	4.84	0.68	1.33	1.35	3.94	6.19
LD	0.51	1.54	10.25	5.36	0.61	8.82	3.93	1.42	1.42	1.35
Sand	0.81	1.81	4.66	2.49	0.39	6.88	0.90	0.80	2.23	1.30
AAT10	2.63	0.49	1.42	0.07	0.38	3.21	1.45	0.87	0.40	7.05
Silt	0.09	0.37	3.69	0.03	0.74	5.38	0.50	0.38	0.48	0.35
Clay	1.94	0.43	0.11	0.40	1.04	0.88	0.74	0.03	0.54	4.18

Abbreviation: AAT=Average annual temperature; AAP=Average annual precipitation; IM=Index of moisture; LF=Landform; GDP=Gross domestic product; AR=Aridity; EL=Elevation; DP=Density of population; LD=Land use; AAT0, ≥ 0 °C annual accumulated temperature; AAT10, ≥ 10 °C Annual accumulated temperature. GLM=Generalized linear model; GBM=Generalized boosted model; GAM=Generalized additive model; CTA=Classification tree analysis; ANN=Artificial neural networks; SRE=Surface range envelope; MARS=Multivariate adaptive regression spline; FDA=Flexible discriminant analysis; RF=Random forest; MaxEnt=Maximum entropy.